Parameter-Dependent Poisson Equations: Tools for Stochastic Approximation in a Markovian Framework*

Algo Carè[1] Balázs Csanád Csáji[2] Balázs Gerencsér[3] László Gerencsér[2] Miklós Rásonyi[3]

***Abstract*— The objective of the present paper is to revisit a key mathematical technology within the theory of stochastic approximation in a Markovian framework, elaborated in much detail in [2]: the existence, uniqueness and smoothness (Lipschitz-continuity) of the solutions of a parameter-dependent Poisson equation. The starting point of our investigation is a relatively new, elegant stability theory for Markov processes developed by Hairer and Mattingly [5]. The current paper provides a transparent analysis of parameter-dependent Poisson equations with convenient conditions. The application of our results for the ODE analysis of stochastic approximation in a Markovian framework is the subject of a forthcoming paper.**

I. Introduction

A beautiful area of systems and control theory is recursive identification, and stochastic adaptive control of stochastic systems. In an abstract mathematical framework [2] [9] the key problem is to solve a non-linear algebraic equation

$$\mathbb{E}\, H(X_n(\theta), \theta) \;=\; 0, \tag{1}$$

where $\theta \in \mathbb{R}^k$ is an unknown, vector-valued parameter of a physical plant or controller, $(X_n(\theta)),\ -\infty < n < +\infty$ is a strictly stationary stochastic process, representing a physical signal affected by θ, and $H(X, \theta)$ is a computable function. The same mathematical framework is applied in other fields such as adaptive signal processing and machine learning.

Our objective is to find the root of (1), denoted by θ^*, via a recursive algorithm based on computable approximations of $H(X_n(\theta), \theta)$. In the case when $H(X_n(\theta), \theta) = h(\theta) + e_n$, where (e_n) is an i.i.d. process, or a martingale difference sequence, we get a classical stochastic approximation process.

An early version of the above problem is presented in the celebrated paper by Ljung [8], in which $(X_n(\theta))$ was assumed to be defined via a linear stochastic system driven by a weakly dependent process.

A renewed interest in recursive estimation in a Markovian framework was sparked by the excellent book of Benveniste, Métivier and Priouret [2] elaborating an extensive mathematical technology for the analysis of these processes. A central tool in their analysis is a complex set of results concerning the parameter-dependent Poisson equation. This is carried out by a specific stability theory for a class of Markov processes, which is off the track of usual methodologies, e.g., Athreya and Ney [1], Nummelin [11], Meyn and Tweedie [10].

The enormous practical value of the estimation problem in a Markovian framework motivates our interest to revisit the theory of [2], and see if their analysis can be simplified or even extended in the light of recent progress in the theory of Markov processes. The starting point of our investigation is a relatively new, elegant stability theory for Markov processes developed by Hairer and Mattingly [5].

The focus of the present paper is the study of the parameter-dependent Poisson equation formulated as

$$(I - P_\theta^*)u_\theta(x) \;=\; f_\theta(x) - h_\theta, \tag{2}$$

where P_θ is the probability transition kernel of the Markov process $(X_n(\theta))$, with $P_\theta^* u_\theta(\cdot)$ denoting the action of P_θ on the unknown function $u_\theta(\cdot)$, and $f_\theta(\cdot)$ is an a priori given function defined on the state-space of the process, finally h_θ denotes the mean value of $f_\theta(\cdot)$ under the assumed unique invariant measure, say μ_θ^*, corresponding to P_θ.

The Poisson equation is a simple and effective tool to study additive functionals on Markov-processes of the form

$$\sum_{n=1}^{N} \big(H(X_n(\theta), \theta) - \mathbb{E}_{\mu_\theta^*} H(X_n(\theta), \theta)\big) \tag{3}$$

via martingale techniques. Proving the Lipschitz continuity of $u_\theta(x)$ w.r.t. θ, and providing useful upper bounds for the Lipschitz constants are vital technical tools for an ODE analysis proposed in [2, Chapter 2, Part II]. The analysis of the Poisson equation takes up more than half of the efforts in proving the basic convergence results in [2], and the verification of their conditions is far from being trivial.

The objective of our project is to revisit the relevant mathematical technologies and outline a hopefully more transparent and flexible analysis within the setup of [5]. The present paper is devoted to the first half of this project, the analysis of the parameter-dependent Poisson equation.

The application of our results for stochastic approximation within a Markovian framework is the subject of a forthcoming paper, in which a combination of the ODE analysis developed in [2] and [4] is to be extended using the results of the current paper. In the end we get the expected rate of convergence for the moments of the estimation error under a convenient set of conditions.

The significance of the topic of the paper is reinforced by the current intense interest in the minimization of functions

*B. Cs. Csáji was supported by the Nat. Res., Dev. and Innov. Office (NKFIH), grant numbers ED_18-2-2018-0006 and KH_17 125698. M. Rásonyi and B. Gerencsér were supported by the NKFIH grant KH 126505 and by the "Lendület" grant LP 2015-6 of the Hungarian Academy of Sciences. B. Gerencsér was also supported by the NKFIH grant PD 121107.

[1]A. Carè is with Dipartimento di Ingegneria dell'Informazione, Università di Brescia, Brescia, Italy, algo.care@unibs.it

[2]L. Gerencsér and B. Cs. Csáji are with MTA SZTAKI: Institute for Computer Science and Control, Hungarian Academy of Sciences, Budapest, Hungary, gerencser@sztaki.hu, csaji@sztaki.hu

[3]B. Gerencsér and M. Rásonyi are with Alfréd Rényi Institute of Mathematics, Budapest, Hungary, gerencser.balazs@renyi.hu, rasonyi@renyi.hu

computed via MCMC [3]. To complement the above historical perspective we should note that the problem goes back to [12], providing results for finite state Markov chains. The extension of these results for more general state-spaces is far from trivial, posing the challenge to choose an appropriate distance of measures.

The structure of the paper is as follows: in Section II we provide a brief introduction to the stability theory for Markov chains developed in [5]. The main results of the paper are stated in Section III, culminating in Theorem 2, proving the Lipschitz continuity of a parameter-dependent Poisson equation. These results are extended in Section IV, in particular, the uniform drift condition, stated as Assumption 1, is significantly relaxed. Our primary objective is to provide a clear, well-motivated presentation of the new concepts and results accompanied by a bird's-eye view on the proofs.

II. A Brief Summary of a New Stability Theory for Markov Chains

Let $(\mathbf{X}, \mathcal{A})$ be a measurable space and $\Theta \subseteq \mathbb{R}^k$ be a domain (i.e., a connected open set). Consider a class of Markov transition kernels $P_\theta(x, A)$, that is for each $\theta \in \Theta$, $x \in \mathbf{X}$, $P_\theta(x, \cdot)$ is a probability measure over $\mathbf{X}$, and for each $A \in \mathcal{A}$, $P_\cdot(\cdot, A)$ is (x, θ)-measurable. Let $(X_n(\theta))$, $n \geq 0$, be a Markov chain with transition kernel P_θ. For any probability measure μ and measurable $\varphi : \mathbf{X} \to \mathbb{R}$ define

$$(P_\theta \mu)(A) = \int_{\mathbf{X}} P_\theta(x, A)\mu(\mathrm{d}x),$$

$$(P_\theta^* \varphi)(x) = \int_{\mathbf{X}} \varphi(y) P_\theta(x, \mathrm{d}y) = \mathbb{E}_\theta\big[\, \varphi(X_1) \mid X_0 = x \,\big],$$

assuming the integral exists. The next condition is motivated by [5], stated there for single Markov chains.

Assumption 1 (Uniform Drift Condition for P_θ): There exists a measurable function $V : \mathbf{X} \to [0, \infty)$ and constants $\gamma \in (0, 1)$ and $K \geq 0$ such that

$$(P_\theta^* V)(x) \leq \gamma V(x) + K, \tag{4}$$

for all $x \in \mathbf{X}$ and $\theta \in \Theta$. Note that $V(x)$ is not θ-dependent.

Remark 1: The drift condition implies that for any probability measure μ such that $\mu(V) := \int_{\mathbf{X}} V(x)\mu(dx) < \infty$,

$$P_\theta \mu(V) \leq \gamma \mu(V) + K. \tag{5}$$

Indeed, integrating (4) with respect to μ we get (5).

As an example, consider a family of linear stochastic systems with state vectors $X_{\theta,n}$:

$$X_{\theta,n+1} = A_\theta X_{\theta,n} + B_\theta U_n,$$

where $\theta \in \Theta$, the matrix A_θ is stable for all $\theta \in \Theta$, and (U_n) is an i.i.d. sequence random vectors such that $\mathbb{E}\,[U_n] = 0$ and $\mathbb{E}\,[U_n U_n^\top] = S$ exists and is finite. Setting $V(x) = x^\top Q x$, where Q is a common symmetric positive definite matrix, it can be easily seen that

$$(P_\theta^* V)(x) = x^\top A_\theta^\top Q A_\theta x + \mathrm{tr}(B_\theta^\top Q B_\theta S).$$

It can be easily seen that the drift condition in the present case is equivalent to $A_\theta^\top Q A_\theta \leq \gamma Q$, with $\gamma < 1$, for all θ, in the sense of the semi-definite ordering.

It may seem too restrictive to assume the existence of a common quadratic Lyapunov function V for all θ. Inspired by alternative conditions that are applicable for this class of processes, Assumption 1 will be relaxed in Section IV.

The next condition is a natural extension of the corresponding assumption of [5] for a parametric family of Markov chains, which itself is a modification of a standard condition in the stability theory of Markov chains [10].

Assumption 2 (Local Minorization): Let $R > 2K/(1-\gamma)$, where γ and K are the constants from Assumption 1, and set $\mathcal{C} = \{x \in \mathbf{X} : V(x) \leq R\}$. There exist a probability measure $\bar{\mu}$ on $\mathbf{X}$ and a constant $\bar{\alpha} \in (0, 1)$ such that, for all $\theta \in \Theta$, all $x \in \mathcal{C}$, and all measurable A,

$$P_\theta(x, A) \geq \bar{\alpha}\bar{\mu}(A).$$

Remark 2 (Interpretation of R): If there exists an invariant measure μ_θ^* such that $\int_{\mathbf{X}} V(x)\mu_\theta^*(\mathrm{d}x) < \infty$, then integrating both sides of inequality (4), we get

$$\int_{\mathbf{X}} V(x)\mu_\theta^*(\mathrm{d}x) \leq \frac{K}{1-\gamma}. \tag{6}$$

Thus, R in Assumption 2 exceeds twice the mean of V w.r.t. any of the invariant measures.

Assumption 2 is a major point of departure from the theory developed in [10], where the "small set" $\mathcal{C}$ is defined in terms of an irreducibility measure ψ such that $\psi(\mathcal{C}) > 0$.

We now introduce a weighted total variation distance between two probability measures μ_1, μ_2, where the weighting is in the form $1+\beta V(\cdot)$, where $\beta > 0$ for which a fine-tuned choice will be needed for the results of [5] to hold.

Definition 1: Let μ_1 and μ_2 be two probability measures on $\mathbf{X}$. Then, define the weighted total variation distance as

$$\rho_\beta(\mu_1, \mu_2) = \int_{\mathbf{X}} (1 + \beta V(x))|\mu_1 - \mu_2|(\mathrm{d}x),$$

where $|\mu_1 - \mu_2|$ is the total variation measure of $(\mu_1 - \mu_2)$.

An equivalent definition of ρ_β can be given by introducing the following norm in the space of $\mathbb{R}$-valued functions on $\mathbf{X}$:

Definition 2: For any function $\varphi : \mathbf{X} \to \mathbb{R}$, set

$$\|\varphi\|_\beta = \sup_x \frac{|\varphi(x)|}{1 + \beta V(x)}. \tag{7}$$

The linear space of real-valued ameasurable functions such that $\|\varphi\|_\beta < \infty$ will be denoted by $\mathcal{L}_V$. Note that $\mathcal{L}_V$ is independent of β. An equivalent definition of ρ_β is:

$$\rho_\beta(\mu_1, \mu_2) = \sup_{\varphi : \|\varphi\|_\beta \leq 1} \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x). \tag{8}$$

Denoting by δ_x the Dirac measure at x, note that, for $x \neq y$, it holds that $\rho_\beta(\delta_x, \delta_y) = 2 + \beta V(x) + \beta V(y)$. This leads to the definition of the following metric on $\mathbf{X}$:

$$d_\beta(x, y) = \begin{cases} 2 + \beta V(x) + \beta V(y) & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases} \tag{9}$$

This may seem to be an unusual metric, assigning a distance at least 2 between any pair of distinct points, but it turns out to be quite useful. Having a metric on $\mathbf{X}$, we can introduce a measure of oscillation for functions $\varphi : \mathbf{X} \to \mathbb{R}$.

Definition 3: For any function $\varphi : \mathbf{X} \to \mathbb{R}$, set

$$|||\varphi|||_\beta = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d_\beta(x, y)}. \tag{10}$$

It is readily seen that $|||\varphi|||_\beta \leq \|\varphi\|_\beta$. Since $|||\varphi|||_\beta$ is invariant w.r.t. translation by any constant $c \in \mathbb{R}$ we also get $|||\varphi|||_\beta \leq \|\varphi + c\|_\beta$. Surprisingly, the infimum, and in fact the minimum, of these upper bounds reproduces $|||\varphi|||_\beta$ as stated in the following lemma proved in [5]:

Lemma 1: $|||\varphi|||_\beta = \min_{c \in \mathbb{R}} \|\varphi + c\|_\beta$.

Definition 4: Let μ_1, μ_2 be two probability measures on $\mathbf{X}$. Then, we define the distance

$$\sigma_\beta(\mu_1, \mu_2) = \sup_{\varphi : |||\varphi|||_\beta \leq 1} \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x). \tag{11}$$

A relatively simple corollary of Lemma 1 is the following:

Corollary 1: For probability measures μ_1, μ_2, we have

$$\sigma_\beta(\mu_1, \mu_2) = \rho_\beta(\mu_1, \mu_2). \tag{12}$$

Remark 3: The metrics $\rho_\beta(\mu_1, \mu_2)$ and $\sigma_\beta(\mu_1, \mu_2)$ depend only on $(\mu_1 - \mu_2)$, therefore they can be expressed by the univariate functions $\rho_\beta(\eta)$ and $\sigma_\beta(\eta)$ defined for signed measures η with $|\eta|(V) < \infty$ and $\eta(\mathbf{X}) = 0$ as

$$\begin{aligned} \sigma_\beta(\eta) &= \int_{\mathbf{X}} (1 + \beta V(x)) |\eta|(\mathrm{d}x) \\ &= \sup_{\varphi : \|\varphi\|_\beta \leq 1} \int_{\mathbf{X}} \varphi(x) \eta(\mathrm{d}x) \\ &= \sup_{\varphi : |||\varphi|||_\beta \leq 1} \int_{\mathbf{X}} \varphi(x) \eta(\mathrm{d}x). \end{aligned} \tag{13}$$

A fundamental result of [5, Theorem 3.1] is as follows:

Proposition 1: Under Assumptions 1 and 2, there exists $\alpha \in (0, 1)$ and $\beta > 0$ such that for all θ and measurable φ,

$$|||P_\theta^* \varphi|||_\beta \leq \alpha |||\varphi|||_\beta. \tag{14}$$

In particular, one can choose $\beta = \bar{\alpha}/(2K)$, and then choose any α such that $\alpha > (1 - \bar{\alpha}/2) \vee \frac{2 + \beta(R\gamma + 2K)}{2 + \beta R}$, where this lower bound can be seen to be strictly less than 1.

Remark 4: Note that with the choice of α as given in Proposition 1 it holds that $1 > \alpha > \gamma$. This indicates that the contraction coefficient α is strictly larger than the contraction coefficient γ postulated by the drift condition.

A corollary of Proposition 1 stated in [5, Theorem 1.3] is:

Proposition 2: Under Assumptions 1 and 2, there exists $\alpha \in (0, 1)$ and $\beta > 0$, such that for all θ,

$$\sigma_\beta(P_\theta \mu_1, P_\theta \mu_2) \leq \alpha \sigma_\beta(\mu_1, \mu_2), \tag{15}$$

for any pair of probability measures μ_1, μ_2 on $\mathbf{X}$.

In what follows, α and β are chosen as indicated in Proposition 1. Using standard arguments one can easily show the following theorem also stated in [5] as Theorem 3.2:

Proposition 3: Under Assumptions 1 and 2 for all θ there is a unique probability measure μ_θ^* on $\mathbf{X}$ such that $\int_{\mathbf{X}} V \mathrm{d}\mu_\theta^* < \infty$ and $P_\theta \mu_\theta^* = \mu_\theta^*$.

Similar results to those of Propositions 2 and 3 are stated in Theorem 14.0.1 [10] under slightly different conditions. In particular, the special choice of the parameter β in the weighting function $1 + \beta V$ is not part of the conditions in [10] at the price that the contraction of the one-step kernel P_θ is not stated. In addition, in [10] it is a priori assumed that the Markov-chain is ψ-irreducible and aperiodic, while in [5] these conditions are circumvented by assuming that the minorization condition holds on a fairly large set.

III. Lipschitz Continuity of the Solution of a θ-Dependent Poisson Equation

In this section we shall consider the Poisson equation

$$(I - P_\theta^*) u_\theta(x) = f_\theta(x) - h_\theta, \tag{16}$$

for $\theta \in \Theta$, where P_θ is given above and $f_\theta : \mathbf{X} \to \mathbb{R}$, $h_\theta = \mu_\theta^*(f_\theta)$, and we look for a solution $u_\theta : \mathbf{X} \to \mathbb{R}$. First, we prove the existence and the uniqueness of the solution for a fixed θ, then we formulate smoothness conditions on the kernel P_θ^*, and the right hand side, f_θ. Using these conditions we prove the Lipschitz continuity of the solution $u_\theta(\cdot)$ in θ. For a start let $\theta \in \Theta$ be fixed.

Theorem 1: Let Assumptions 1 and 2 hold. Let f be a measurable function $\mathbf{X} \to \mathbb{R}$ such that $|||f|||_\beta < \infty$ and let $P = P_\theta$ for some fixed θ, with invariant measure $\mu^* = \mu_\theta^*$. Let $h = \mu^*(f)$. Then, the Poisson equation

$$(I - P^*) u(x) = f(x) - h \tag{17}$$

has a unique solution $u(\cdot)$ up to an additive constant. Henceforth, we shall consider the particular solution

$$u(x) = \sum_{n=0}^{\infty} (P^{*n} f(x) - h), \tag{18}$$

which is well-defined, in fact the right hand side is absolute convergent, and in addition $\mu^*(u) = 0$. Furthermore,

$$|u(x)| \leq |||f|||_\beta K(x), \tag{19}$$

where $K(x) := \frac{1}{1-\alpha} \left(2 + \beta V(x) + \beta \frac{K}{1-\gamma} \right)$, also implying $\|u\|_\beta < \infty$.

Outline of the proof: It is immediate to check that (17) is formally satisfied by u. To show that u is well-defined, use:

$$\left| \int_{\mathbf{X}} \varphi(x)(\mu_1 - \mu_2)(\mathrm{d}x) \right| \leq |||\varphi|||_\beta \sigma_\beta(\mu_1, \mu_2). \tag{20}$$

For the nth term of the right hand side of (18), we have:

$$\begin{aligned} \frac{1}{|||f|||_\beta} |P^{*n} f(x) - \mu^*(f)| &= \frac{1}{|||f|||_\beta} |(P^n \delta_x - \mu^*)(f)| \\ &= \frac{1}{|||f|||_\beta} \left| \int_{\mathbf{X}} f(y)(P^n \delta_x - P^n \mu^*)(\mathrm{d}y) \right|. \end{aligned}$$

We can bound the right hand side by

$$\sigma_\beta(P^n\delta_x, P^n\mu^*) \le \alpha^n \sup_{\varphi:\|\varphi\|_\beta\le 1} \int_{\mathbf{X}} \varphi(x)(\delta_x - \mu^*)(\mathrm{d}x).$$

We conclude that the series $\sum_{n=0}^{\infty}(P^{*n}f(x) - h)$ is absolutely convergent, so $u(x)$ is well-defined and satisfies the desired upper bound. It is readily seen that

$$\int_{\mathbf{X}} u(x)\mu^*(\mathrm{d}x) = 0. \tag{21}$$

The uniqueness follows directly from Proposition 1.

Now we consider a parametric family of kernels (P_θ) and that of functions (f_θ) for $\theta \in \Theta$, and impose appropriate smoothness conditions for them in the context of [5].

Assumption 3: There exists a constant L_P such that for every $\theta, \theta' \in \Theta$ and $x \in \mathbf{X}$ it holds that

$$\sigma_\beta(P_\theta\delta_x, P_{\theta'}\delta_x) \le L_P|\theta - \theta'|(1 + \beta V(x)). \tag{22}$$

It is easy to show that, under a relaxed drift condition defined by Assumption 1 without assuming $\gamma < 1$, and under Assumption 3, we have for every $\theta, \theta' \in \Theta$ and every probability measure μ such that $\mu(V) < \infty$, the inequality

$$\sigma_\beta(P_\theta\mu, P_{\theta'}\mu) \le L_P|\theta - \theta'|\mu(1 + \beta V). \tag{23}$$

The above observation is easily extended from probability measures to signed measures η such that $|\eta|(V) < \infty$.

The class of functions $\{f_\theta : \mathbf{X} \to \mathbb{R} \mid \theta \in \Theta\}$ is characterized by the following assumption:

Assumption 4: We have $K_f := \sup_{\theta\in\Theta} |||f_\theta|||_\beta < \infty$, and there exists a constant L_f such that, for all θ, θ', it holds that

$$\|f_\theta - f_{\theta'}\|_\beta \le L_f|\theta - \theta'|. \tag{24}$$

The main result of the paper is as follows.

Theorem 2: Let Assumptions 1, 2, 3 and 4 hold, and consider the parameter-dependent Poisson equation

$$(I - P_\theta^*)u_\theta(x) = f_\theta(x) - h_\theta, \tag{25}$$

where $h_\theta = \mu_\theta^*(f_\theta)$. Then, h_θ is Lipschitz continuous in θ:

$$|h_\theta - h_{\theta'}| \le L_h|\theta - \theta'|, \tag{26}$$

and the family of solutions $u_\theta(x) = \sum_{n=0}^{\infty}(P_\theta^{*n}f_\theta(x) - h_\theta)$, ensured by Theorem 1, is Lipschitz continuous in θ:

$$|u_\theta(x) - u_{\theta'}(x)| \le L_u\,(1 + \beta V(x))\,|\theta - \theta'|,$$

where the constant L_u is independent of x. Note that this also implies $\|u_\theta - u_{\theta'}\|_\beta \le L_u|\theta - \theta'|$.

Outline of the proof: Consider the extended parametric family of Poisson-equations, where P^* and f are independently parametrized, with the notation $h_{\theta,\psi} = \mu_\theta^*(f_\psi)$,

$$(I - P_\theta^*)u_{\theta,\psi}(x) = f_\psi(x) - h_{\theta,\psi}, \tag{27}$$

First, we prove that $h_{\theta,\psi}$ is Lipschitz-continuous in θ and ψ. Since $h_\theta = \mu_\theta^*(f_\theta) = h_{\theta,\theta}$, the Lipschitz-continuity of h_θ, stated in (26) then follows. We can write

$$|h_{\theta,\psi} - h_{\theta,\psi'}| = \lim_{n\to\infty} |P_\theta^{*n}f_\psi(x) - P_\theta^{*n}f_{\psi'}(x)|, \tag{28}$$

$$|h_{\theta,\psi} - h_{\theta',\psi}| = \lim_{n\to\infty} |P_\theta^{*n}f_\psi(x) - P_{\theta'}^{*n}f_\psi(x)|. \tag{29}$$

We can bound the right hand side of (28) as follows:

$$\begin{aligned}|P_\theta^{*n}f_\psi(x) - P_\theta^{*n}f_{\psi'}(x)| &\le (P_\theta^{*n}|f_\psi - f_{\psi'}|)(x)\\ &= (P_\theta^n\delta_x)\,|f_\psi - f_{\psi'}|.\end{aligned} \tag{30}$$

Using the Lipschitz continuity of f as given by Assumption 4 and the drift condition Assumption 1, we finally get

$$\limsup_{n\to\infty} |P_\theta^{*n}f_\psi(x) - P_\theta^{*n}f_{\psi'}(x)| \le L_f|\psi - \psi'|\left[1 + \beta\frac{K}{1-\gamma}\right].$$

To continue the proof of the we will have to establish the Lipschitz-continuity of the powers of the kernel P_θ^n together with an upper bound for the Lipschitz constants. We can show that for any probability measure μ with $\mu(V) < \infty$,

$$\sigma_\beta(P_\theta^n\mu, P_{\theta'}^n\mu) \le L_P|\theta - \theta'|\left(L_P' + \frac{\alpha^n}{\alpha - \gamma}\beta\mu(V)\right), \tag{31}$$

where L_P' is determined by the constants showing up in the assumptions for P_θ. The proof is obtained by using a kind of telescopic inequality.

A direct corollary is that for measurable functions φ with $|||\varphi|||_\beta < \infty$ it holds that $|P_\theta^{*n}\varphi(x) - P_{\theta'}^{*n}\varphi(x)|$ is bounded from above by

$$|||\varphi|||_\beta L_P|\theta - \theta'|\left(L_P' + \frac{\alpha^n}{\alpha - \gamma}\beta V(x)\right). \tag{32}$$

From (31) above we immediately get the Lipschitz-continuity of the invariant measures with $L_P'' = L_P L_P'$:

$$\sigma_\beta(\mu_\theta^*, \mu_{\theta'}^*) \le L_P''|\theta - \theta'|. \tag{33}$$

Inequality (31) has an effective extension for signed measures η satisfying the additional condition $\eta(\mathbf{X}) = 0$:

Lemma 2: Assume that Assumptions 1, 2, and 3 hold. Then for every $\theta, \theta' \in \Theta$ and every signed measure η such that $|\eta|(V) < \infty$ and $\eta(\mathbf{X}) = 0$, we have

$$\sigma_\beta(P_\theta^n\eta, P_{\theta'}^n\eta) \le L_P|\theta - \theta'|n\alpha^{n-1}\,|\eta|(1 + \beta V). \tag{34}$$

Returning to the right hand side of (29) we use the upper bound (32) with $\varphi = f_\psi$ and let n go to infinity:

$$\limsup_{n\to\infty} |P_\theta^{*n}f_\psi(x) - P_{\theta'}^{*n}f_\psi(x)| \le |||f_\psi|||_\beta L_P''|\theta - \theta'|. \tag{35}$$

Next, we consider the Lipschitz continuity of the doubly-parametrized particular solution

$$u_{\theta,\psi}(x) = \sum_{n=0}^{\infty}(P_\theta^{*n}f_\psi(x) - h_{\theta,\psi}). \tag{36}$$

The critical point is to show that $u_{\theta,\psi}(x)$ is Lipschitz-continuous in θ. Consider the measure in the n-th term:

$$[P_\theta^n\,(\delta_x - \mu_\theta^*) - P_{\theta'}^n\,(\delta_x - \mu_\theta^*)] + [P_{\theta'}^n\,(\mu_{\theta'}^* - \mu_\theta^*)].$$

The second term of the right hand side can be readily handled by (33), while the first term can be dealt with using Lemma 2 setting $\eta = \delta_x - \mu_\theta^*$. The rest of the proof is analogous to the proof of Theorem 1.

IV. Relaxations of the Uniform Drift Condition

A delicate condition of Propositions 1-3 is Assumption 1, requiring the existence of a common Lyapunov function. This requirement may be too restrictive even in the case of linear stochastic systems as discussed in Section II. However, assuming that (A_θ), $\theta \in \Theta$ is a compact set of stable matrices we can find a positive integer r such that $\|A_\theta^r\| \leq \gamma_r < 1$ for all $\theta \in \Theta$. This example motivates the following relaxation of the drift condition, given as Assumption 1:

Assumption 5 (Uniform Drift Condition for P_θ^r): There exists a positive integer r, a measurable function $V : \mathbf{X} \to [0, \infty)$ and constants $\gamma_r \in (0, 1)$ and $K_r \geq 0$ such that for all $\theta \in \Theta$ and $x \in \mathbf{X}$, we have

$$(P_\theta^{*r} V)(x) \leq \gamma_r V(x) + K_r, \tag{37}$$

and the following uniform one-step growth condition holds:

$$(P_\theta^* V)(x) \leq \gamma_1 V(x) + K_1, \tag{38}$$

where we can and will assume that $\gamma_1 > 1$ and $K_1 \geq 0$.

Note that (38) implies that for any $\beta > 0$ there exist $C' > 0$ such that for any function $\varphi \in \mathcal{L}_V$ we have

$$|||P_\theta^* \varphi|||_\beta \leq \alpha' |||\varphi|||_\beta, \tag{39}$$

for all θ with $\alpha' = \max(1 + \beta K_1, \gamma_1)$. From here, repeating the arguments leading to Proposition 2, we get:

Lemma 3: Assume (38), then for any pair of probability measures μ_1, μ_2 on $\mathbf{X}$ such that $\mu_1(V), \mu_2(V) < \infty$ and any $\beta > 0$, we have for all θ,

$$\sigma_\beta(P_\theta \mu_1, P_\theta \mu_2) \leq \alpha' \sigma_\beta(\mu_1, \mu_2), \tag{40}$$

Assumption 6 (Uniform Local Minorization for P_θ^r): Let $R_r > 2K_r/(1 - \gamma_r)$ where γ_r and K_r are the constants from Assumption 5 and $\mathcal{C}_r = \{x \in \mathbf{X} : V(x) \leq R_r\}$. There exist a probability measure $\bar{\mu}_r$ and a constant $\bar{\alpha}_r \in (0, 1)$ such that for all $\theta \in \Theta$, $x \in \mathcal{C}_r$ and measurable A it holds

$$P_\theta^r(x, A) \geq \bar{\alpha}_r \bar{\mu}_r(A). \tag{41}$$

The main results cited in Section II can be extended, with minor modifications, assuming the above relaxed conditions. For now we fix any $\theta \in \Theta$ and write $P_\theta = P$. Proposition 1 can be restated as follows:

Theorem 3: Under Assumptions 5 and 6 there exist $\alpha \in (0, 1), \beta > 0$ and $C > 0$ such that for any measurable φ and $n > 0$ we have

$$|||P^{*n} \varphi|||_\beta \leq C \alpha^n |||\varphi|||_\beta,$$

where we can choose $\beta = \beta_r$, given by Proposition 1 applied to P^r, $\alpha = \alpha_r^{1/r}$ with some $C > 0$.

Proof: By Proposition 1 there exist $\beta = \beta_r > 0$, and $\alpha_r \in (0, 1)$ such that $|||P^{*r} \varphi|||_\beta \leq \alpha_r |||\varphi|||_\beta$, implying for any positive integer m

$$|||P^{*rm} \varphi|||_\beta \leq \alpha_r^m |||\varphi|||_\beta. \tag{42}$$

For a general positive integer n write $n = rm + k$ with $0 \leq k \leq r - 1$ to get

$$|||P^{*n} \varphi|||_\beta \leq \alpha_r^m \left|\left|\left| P^{*k} \varphi \right|\right|\right|_\beta. \tag{43}$$

To complete the proof apply (39) and obtain

$$|||P^{*n} \varphi|||_\beta \leq \alpha_r^m (C')^{r-1} |||\varphi|||_\beta. \tag{44}$$

Now $m = (n - k)/r > n/r - 1$, hence $\alpha_r^m < \alpha_r^{n/r} \alpha_r^{-1}$, and thus the claim follows. ■

Proposition 2 takes now the following modified form:

Theorem 4: Under Assumptions 5 and 6 there exist $\alpha \in (0, 1), \beta > 0$ and $C > 0$ such that for any $n > 0$,

$$\sigma_\beta(P^n \mu_1, P^n \mu_2) \leq C \alpha^n \sigma_\beta(\mu_1, \mu_2), \tag{45}$$

for every pair of probability measures μ_1, μ_2 on $\mathbf{X}$, where α and C are given in Theorem 3.

Finally, we have the following extension of Proposition 3:

Theorem 5: Under Assumptions 5 and 6 there exists a unique probability measure μ^* on $\mathbf{X}$ such that $\int_{\mathbf{X}} V \mathrm{d}\mu^* < \infty$ and $P\mu^* = \mu^*$. Denoting the unique invariant probability measure for P^r by μ_r^* we have $\mu^* = \mu_r^*$.

Proof: Let μ_r^* be the unique invariant probability measure for P^r the existence of which is ensured by Proposition 3. Then $\int_{\mathbf{X}} V \mathrm{d}\mu_r^* < \infty$ implies $\int_{\mathbf{X}} V \mathrm{d}(P^k \mu_r^*) < \infty$ for any $k > 0$ by the one-step growth condition, see (39). It follows that the probability measure μ defined by

$$\mu = \frac{1}{r}(I + P + \ldots P^{r-1}) \mu_r^*$$

also satisfies $\int_{\mathbf{X}} V \mathrm{d}\mu < \infty$, and it is readily seen that it is invariant for P. Since any probability measure invariant for P is also invariant for P^r, we have $\mu = \mu_r^*$. The uniqueness of an invariant probability measure for P follows by noting once again if μ' is invariant for P then it is also invariant for P^r, and hence we must have $\mu' = \mu_r^*$. ■

The main results of Section III can now be extended, with minor modifications, assuming the above relaxed conditions. For the extension of Theorem 1 we fix once again any $\theta \in \Theta$ and write $P_\theta = P$:

Theorem 6: Assume that the kernel P^r satsifies Assumptions 5 and 6. Let $\beta > 0$ be as given in Proposition 1 w.r.t. the kernel P^r. Let f be a measurable function such that $|||f|||_\beta < \infty$. Let μ^* denote the unique invariant probability measure of P, and $h = \mu^*(f)$. Then, the Poisson equation

$$(I - P^*) u(x) = f(x) - h \tag{46}$$

has a unique solution u up to additive constants, and considering the particular solution u with $\mu^*(u) = 0$, we have

$$|u(x)| \leq K(1 + \beta V(x)) |||f|||_\beta \tag{47}$$

for some constant $K > 0$ depending only on the constants appearing in Assumptions 5 and 6.

Outline of the proof: The starting point is the Poisson equation for P^{*r}, noting that $h = \mu^*(f) = \mu_r^*(f)$,

$$(I - P^{*r})v(x) = f(x) - h. \tag{48}$$

Consider the particular solution

$$v(x) = \sum_{n=0}^{\infty}(P^{*nr}f(x) - h). \tag{49}$$

It is easy to see that

$$u(x) := (I + P^* + \ldots + P^{*(r-1)})v(x) \tag{50}$$

is a solution of (46) and satisfies (47). Considering the uniqueness of the solution, for the difference of two solutions Δu we have $P^*\Delta u(x) = 0$, for all x. Then applying $r-1$ times P^* we get $P^{*r}\Delta u(x) = 0$, for all x, and thus by Theorem 1 we conclude that Δu is a constant function.

A straightforward extension of Theorem 2 is the following:

Theorem 7: Assume that the kernels (P_θ^r) satsify Assumptions 5 and 6. Let $\beta > 0$ be as given in Proposition 1 w.r.t. the kernel (P_θ^r). Assume (P_θ) also satisfy Assumption 3. Finally, let (f_θ) be a family of measurable functions $\mathbf{X} \to \mathbb{R}$ such that Assumption 4 holds. Let μ_θ^* denote the unique invariant probability measure of P_θ, and let $h_\theta = \mu_\theta^*(f_\theta)$. Consider the parameter-dependent Poisson equation

$$(I - P_\theta^*)u_\theta(x) = f_\theta(x) - h_\theta. \tag{51}$$

Then, h_θ is Lipschitz continuous in θ:

$$|h_\theta - h_{\theta'}| \le L_h|\theta - \theta'|, \tag{52}$$

and the particular solution $u_\theta(x) = \sum_{n=0}^{\infty}(P_\theta^{*n}f_\theta(x) - h_\theta)$ is well-defined for all θ, and Lipschitz continuous in θ,

$$|u_\theta(x) - u_{\theta'}(x)| \le L_u|\theta - \theta'|(1 + \beta V(x)), \tag{53}$$

where the constants L_h and L_u are independent of x.

Outline of the proof: First we prove that $h_\theta = \mu_{\theta,r}^*(f_\theta)$ is Lipschitz-continuous referring to Theorem 2 with P_θ^r replacing P_θ. For this we will have to verify Assumption 3 (with P_θ^r replacing P_θ). This is done by extending (31) assuming only the validity of Assumption 3 for P_θ and the uniform one-step growth condition, see Assumption 5. We get for any pair $\theta, \theta' \in \Theta$, for any probability measure μ such that $\mu(V) < \infty$ and for any $n > 0$ we have

$$\sigma_\beta(P_\theta^n\mu, P_{\theta'}^n\mu) \le L_P''|\theta - \theta'|(\alpha')^n\left(1 + \beta\mu(V)\right), \tag{54}$$

choosing $\alpha' > \gamma_1$, with L_P'' depending only on n and the constants appearing in the conditions of the theorem.

It follows, in view of Theorem 2, that the particular solution of the Poisson equation

$$(I - P_\theta^{*r})v_\theta(x) = f_\theta(x) - h_\theta \tag{55}$$

given by $v_\theta(x) = \sum_{n=0}^{\infty}P_\theta^{*nr}(f_\theta(x) - h_\theta)$ is Lipschitz-continuous and satisfies

$$|v_\theta(x) - v_{\theta'}(x)| \le L_v|\theta - \theta'|(1 + \beta V(x)). \tag{56}$$

Recalling that $(P_\theta^{*m}f_\theta)(x) = P_\theta^m\delta_x(f)$, using (54) it is readily seen that the solution of (51) defined by

$$u_\theta(x) := (I + P_\theta^* + \ldots + P_\theta^{*(r-1)})v_\theta(x) \tag{57}$$

is Lipschitz continuous in θ, and due to the one-step growth condition it satisfies (53), completing the proof.

V. Discussion

The verification of Assumption 5 may seem to be too demanding. We propose a simple alternative criterion:

Assumption 7 (Individual Drift Conditions): There exists a family of measurable functions $V_\theta : \mathbf{X} \to [0, \infty)$ and constants $\gamma \in (0, 1)$ and $K \ge 0$ such that for all x and θ

$$(P_\theta^*V_\theta)(x) \le \gamma V_\theta(x) + K, \tag{58}$$

moreover, there exists a measurable $V : \mathbf{X} \to [0, \infty)$ and constants a, b, c, d with $a, c > 0$, such that

$$aV(x) + b \le V_\theta(x) \le cV(x) + d. \tag{59}$$

Under Assumption 7, for any sufficiently large r Assumption 5 is satisfied with the function V. It is also easily seen that Theorem 7 remains valid under conditions imposed on the one-step kernels (P_θ), namely Assumptions 7 and 2.

A possible alternative set of conditions under which the problems of the paper may be worth studying is provided by the theory developed in [10], extended in later works, such as [6] and [7]. However, the extension of Assumption 3 on the Lipschitz-continuity of P_θ, so that the Lipschitz-continuity of $(I - P_\theta)^{-1}$ is implied, does not seem obvious.

References

[1] Krishna B. Athreya and Peter Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245:493–501, 1978.

[2] Albert Benveniste, Michel Métivier, and Pierre Priouret. *Adaptive algorithms and stochastic approximations*. Springer, 2nd edition, 1990.

[3] Nicolas Brosse, Alain Durmus, Sean Meyn, Éric Moulines, and Anand Radhakrishnan. Diffusion approximations and control variates for MCMC. *arXiv:1808.01665*, 2018.

[4] László Gerencsér. Rate of convergence of recursive estimators. *SIAM Journal on Control and Optimization*, 30(5):1200–1227, 1992.

[5] Martin Hairer and Jonathan C. Mattingly. Yet another look at Harris ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117. Springer, 2011.

[6] Ioannis Kontoyiannis and Sean Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *The Annals of Applied Probability*, 13(1):304–362, 2003.

[7] Ioannis Kontoyiannis and Sean Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electronic Journal of Probability*, 10:61–123, 2005.

[8] Lennart Ljung. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22(4):551–575, 1977.

[9] Lennart Ljung and Torsten Söderström. *Theory and practice of recursive identification*. MIT Press, 1983.

[10] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag London, 2009.

[11] Esa Nummelin. A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43(4):309–318, 1978.

[12] Paul J. Schweitzer. Perturbation theory and finite Markov chains. *Journal of Applied Probability*, 5(2):401–413, 1968.